A MATHEMATICAL TREATMENT OF THE SPHERICAL STEREOLOGY

Charles Hagwood

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards
and Technology
Center for Computing & Applied
Mathematics
Statistical Engineering Division
Gaithersburg, MD 20899

NIST

A MATHEMATICAL TREATMENT OF THE SPHERICAL STEREOLOGY

Charles Hagwood

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards
and Technology
Center for Computing & Applied
Mathematics
Statistical Engineering Division
Gaithersburg, MD 20899

July 1990

Contents

1 Introduction

In R^3 a collection of arbitrarily shaped small particles are embedded in an opaque medium, and the problem is to study their size distribution. Due to physical impediments one cannot observe the particles directly, but planar probes can be used to retrieve information on the size distribution. More specifically, a test plane is randomly chosen from among all planes parallel to the z-axis and intersecting the medium. The intersection of the 3-dimensional particles with the test plane produces 2-dimensional figures. The size distribution of the 2-dimensional figures are used to make inference about the 3-dimensional size distribution of particles in the medium. Although this seems an impossible task, we will show that for particles of certain shapes, an exact relationship exists. Problems of this form occur in biology, materials science, ceramics, mineralogy, metallurgy etc.

First we make two assumptions

A1. The test planes are chosen uniformly along the z-axis i.e. the probability of choosing a test plane in any interval of width δz is the same as for any other interval of length δz

A2. The centers of mass of the particles are distributed according to a Poisson process with parameter, λ, where λ denotes the mean number of centers per unit volume.

If A2 holds with λ sufficiently small, then for small particles we may assume that the particles are so dilutely distributed that their positions and orientations are independent of one another. Also, the assumption implies that there exists a finite number of particles in any region of finite volume. The above set-up produces a marked point process in R^3, (C_λ, R) where $C_\lambda, \lambda > 0$ is the Poisson process describing the centers of mass and R is a collection of parameters describing the particle.
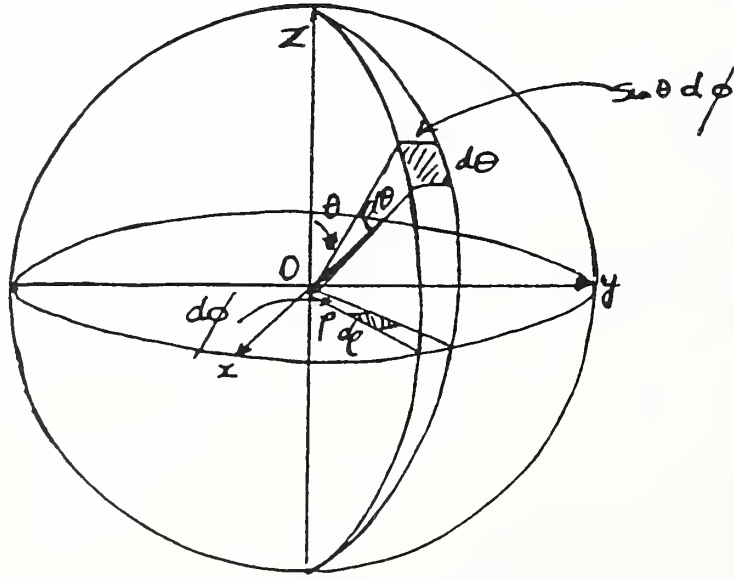
2 Fundamental relation of stereology

One quantity of theoretical and practical importance in describing the size distribution of particles, is N_V, the number of particles per unit volume. In this section it is shown that the observed number of particles per unit area, N_A, is related to N_V via

$$\bar{N}_A = N_V \bar{H}$$

This relationship is called the fundamental relationship of stereology, and it holds for particles of arbitrary shape, and size. $\bar{H}$ is the average distance between tangent planes over all particles in the structure, this is clarified below.

One readily sees that by cutting the medium with a random plane and observing the 2-D images is in general ambiguous, since different orientations of the particles about their centers of mass produce different 2-D images. To correct for this we define a uniform distribution of orientations as follows. Vaguely one can define for oneself the meaning to the statement, the orientation of a particle about its center of mass. Specifically, what one means by this statement is: Let (x_0, y_0, z_0) denote the center of mass of the particle. Place an x, y, z coordinate axis at the (x_0, y_0, z_0). By the orientations of the particle about its center of mass, we mean the different positions the particle takes as the axis is rotated through the angles θ and ϕ ranging over $0 \le \phi \le \pi$, $0 \le \phi \le 2\pi$. See the figure below. The uniform distribution is given by.

$$f(\theta, \phi) = sin(\phi)d\theta d\phi$$

The fundamental relation of stereology is now derived. Let V be a region of finite volume, then by A2 we may assume a finite number of particles in the region. Let

n_0 = the number of particles(centers) in the region V

$|V|$ = volume of V

N_V = number of particles per unit volume

n_p = the number of particles that intersect an arbitrary test plane A

N_A = the number of particles per unit test area

Then

$$N_V = \frac{n_0}{|V|}$$
$$N_A = \frac{n_p}{|A|}$$

Note that n_p and N_A depend on the orientations of the particles about their centers of mass. Let (θ_i, ϕ_i) be the orientation vector for the ith particle $i = 1, \ldots, n_0$. We placed a uniform distribution on the set of orientations so (θ_i, ϕ_i) is a random variable on the sphere. Therefore we write

$$n_p = n_p((\theta_1, \phi_1), \ldots, (\theta_{n_0}, \phi_{n_0})$$

as a random variable depending on the independent variables (θ_i, ϕ_i).

The problem is to find a relationship between N_V and N_A. Suppose the medium is a cube with sides of length l. Then

$$V = l^3 \quad \text{and} \quad A = l^2$$
$$N_V = \frac{n_0}{l^3} \quad \text{and} \quad N_A = \frac{n_p}{l^2}$$

Consider first just one fixed particle with orientation (θ, ϕ) about its center of mass. Let $H(\theta, \phi)$ be the distance between parallel horizontal tangent planes to the particle, as shown below.

3

It is easily seen that under the probability distribution given by A1 the probability that a randomly chosen plane will intersect the particle is given by
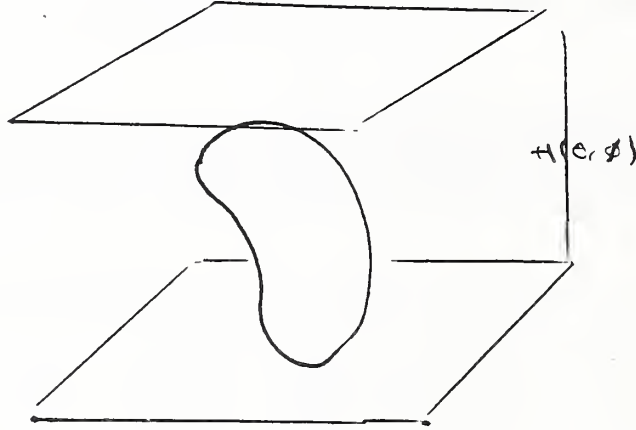
$$\Pr(\text{test plane intersects particle} \mid \theta, \phi) \ = \ \frac{H(\theta,\phi)}{l})$$

Now assume that there are t types of particles with n_1 of type 1, n_2 of type 2,...,n_t of type t. Let

p_{ij} denotes the probability a randomly chosen test plane hits the jth particle of type i given its orientation (θ_{ij}, ϕ_{ij})

then

$$p_{ij} = \frac{H(\theta_{ij}, \phi_{ij})}{l}$$

$j = 1, \ldots, n_i, i = 1, \ldots, t$ where $H_i(\theta, \phi)$ denotes the distance between tangent planes for particle i given orientation (θ, ϕ). Let

$$I_{ij} = \begin{cases} 1 & \text{if the ith particle of type j is hit by the test plane} \\ 0 & \text{otherwise} \end{cases}$$

4

Then, $I_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, t$ from an array of independent Bernoulli random variables with means p_{ij}. That is, a particle is either on the test plane or it is not, independent of the other particles. We have

$$n_p = \sum_{i=1}^{t} \sum_{j=1}^{n_i} I_{ij}$$

and

$$
\begin{aligned}
E[n_p \mid (\theta_{ij}, \phi_{ij}, i, j] &= \sum_{i=1}^{t} \sum_{j=1}^{n_i} p_{ij} \\
&= \frac{1}{l} \sum_{i=1}^{t} \sum_{j=1}^{n_i} H(\theta_{ij}, \phi_{ij}) \\
&= \frac{n_0}{l} \sum_{i=1}^{t} \frac{1}{n_0} \sum_{j=1}^{n_i} H(\theta_{ij}, \phi_{ij})
\end{aligned}
$$

Averaging over all orientations gives

$$
\begin{aligned}
E[n_p] &= \frac{n_0}{l} \sum_{i=1}^{t} \frac{1}{n_0} \sum_{j=1}^{n_j} \int_0^{2\pi} \int_0^{\pi} H_i(\theta_{ij}, \phi_{ij}) f(\theta_{ij}, \phi_{ij}) d\theta_{ij} d\phi_{ij} \\
&= \frac{n_0}{l} \sum_{i=1}^{t} \frac{n_i}{n_0} \int_0^{2\pi} \int_0^{\pi} H_i(\theta, \phi) f(\theta, \phi) d\theta d\phi \\
&= \frac{n_0}{l} \overline{H}
\end{aligned}
$$

Thus,

$$E[n_p] = \frac{n_0}{l} \overline{H}$$

where $\overline{H} = \sum_{i=1}^{n_0} \frac{n_i}{n_0} \overline{H_i}(\theta, \phi)$ and $\overline{H_i} = \int_0^{2\pi} \int_0^{\pi} H_i(\theta, \phi) f(\theta, \phi) d\theta d\phi$ denotes the average distance between parallel tangent planes for the ith particle. The average is with respect to orientation. Since, $\frac{n_i}{n_0} i = 1, \ldots, t$ forms a distribution, then $\overline{H}$ denotes a grand average distance over all particles. If we divide both sides of the equation

$$E[n_p] = \frac{n_0}{l} \overline{H}$$

5

by l^2 gives

$$\frac{E[n_p]}{l^2} = \frac{n_0\overline{H}}{l^3}$$

since $N_V = \frac{n_0}{l^3}$ and $N_A = \frac{n_p}{l^2}$ it follows that

$$E[N_A] = E[\frac{n_p}{l^2} = \frac{n_0}{l^3}\overline{H} = N_V\overline{H}$$

or

$$E[N_A] = N_V\overline{H}$$

This is the **fundamental relation of stereology**, which is usually written in the form

$$\overline{N}_A = N_V\overline{H}$$

which says that the average density of particles on a randomly selected test plane is proportional to the density of particles in the medium, with proportionality constant , $\overline{H}$.

If the particles are spheres, then all orientations of a sphere about its center of mass are the same and so $H_i = 2R_i$, the radius of the sphere. If there are t types of particles, then they are described by t distinct radii, $R_i, i = 1, \ldots, t$, and we have from above

$$E[N_A] = 2\overline{R}N_V$$

where $\overline{R} = \sum_{i=1}^{n_0} \frac{n_i}{n_0}2R_i$, the average spherical diameter. The expectation above is taken solely with respect to Lebesgue measure. Furthermore, we note the following. We may write

$$n_p = Y_1 + \cdots + Y_t$$

where Y_i denotes the number of particles of type i on a randomly selected test plane. The Y_i are independent binomial random variables with parameters p_i and n_i By the law of large numbers

$$\lim_{l\to\infty} | \frac{n_p}{l^2} - 2\bar{R}N_V] |= \lim_{l\to\infty} | E[N_A] - 2\overline{R}N_V |= 0$$

6

provided $\min\{n_i\} \to \infty$ as $l \to \infty$.

3 Spheres as Particles

If the particles are spheres then they can be described uniquely by their radii. Let $F(r)$, and $f(r)$, $r \geq 0$ denote the unknown radii distributionand density, respectively of the spheres, i.e.

$$F(r) = Pr\{R_s \leq r\}$$

where R_s denotes the radius of a randomly chosen sphere. Let $z = z_0$ be a plane parallel to the z axis, then obviously, this plane induces a distribution of circular radii given by the intersection of the spheres with the plane. Let $G(r)$ and $g(r)$ be the profile size distribution and density,respectively, i.e.

$$G(r) = Pr\{R_c \leq r\}$$

where R_c denotes the radius of a randomly selected circle on a randomly selected test plane.The orginal problem can be restated simply as: find a relationship between $F(r)$ and $G(r)$, assuming that an estimate of $G(r)$ exists.

Consider the sphere

$$x^2 + y^2 + z^2 = R^2$$

and let $z = z_0$, we get a test plane parallel to the $z-$axis. If $z = z_0$ intersects the sphere, we get the circle

$$\begin{aligned} x^2 + y^2 + z_0^2 &= R^2 \\ x^2 + y^2 &= R^2 - z_0^2 \end{aligned}$$

This induces a function

$$r(z_0) = \sqrt{R^2 - z_0^2}$$

which denotes the radius of the circular profile cut out by the plane $z = z_0$, $-R \leq z_0 \leq R$. Therefore

$$r(Z) = \sqrt{R^2 - Z^2}$$

7

is a random variable, where Z is the Z-coordinate of uniformly selected test plane .

Consider just that portion of R^3 in the cube centered at zero and with sides of length l. Suppose we uniformly select a test plane with $-l \leq Z \leq l$. Let R_s denote the radius defined above, then

$$
\begin{aligned}
T_R(r \mid R_s) &= P(r(Z) \leq r \mid Z \text{ intersects the sphere}) \\
&= P(\sqrt{R^2 - Z^2} \leq r \mid Z \text{ intersects the sphere}) \\
&= P(R^2 - Z^2 \leq r^2 \mid Z \cap \text{ sphere}) \\
&= P(Z^2 \geq R^2 - r^2 \mid Z \cap \text{ sphere}) \\
&= 1 - P(Z^2 \leq R^2 - r^2 \mid Z \cap \text{ sphere}) \\
&= 1 - P(-\sqrt{R^2 - r^2} \leq Z \leq \sqrt{R^2 - r^2} \mid Z \cap \text{ sphere}) \\
&= 1 - \frac{2\sqrt{R^2 - r^2}}{2R}
\end{aligned}
$$

with density

$$
t_R(r) = \frac{r}{R\sqrt{R^2 - r^2}}
$$

Observe that when R^3 is cut by the cube of volume l^3 we get a random sample $R_1^s, \ldots, R_{n_0}^s$ of spherical radii R^s. This follows since the centers of the spheres are distributed in R^3 according to a Poisson process. And, when this cube is cut by a test plane, Z, we get a random sample from R^c, namely, $R_1^c, \ldots, R_{n_p}^c$, where $\{R_1^c, \ldots, R_{n_p}^c\} = \{R_1^s, \ldots, R_{n_0}^s\} \cap Z$.

Suppose that the random variable R^s lies between $(0, R_{\max}]$. Then also, R^c has support $(0, R_{\max}]$. Choose $t > 0$ and divide $(0, R_{\max}]$ into t subintervals, I_j, each of equal width $\triangle r = \frac{R_{max}}{t}$. Let $x_i = i \triangle r, i = 1, \ldots, t$ then

$$
I_1 = (0, x_1] \quad , I_2 = (x_1, x_2] \quad \ldots, I_t = (x_{t-1}, x_t]
$$

denote the subintervals. Let

$R_{11}^c, \ldots, R_{1m_1}^c$ denote the circles given by the intersection of Z with spheres with $R^s \in I_1$

$R_{21}^c, \ldots, R_{2m_2}^c$ denote the circles given by the intersection of Z with $R^s \in I_2$

8

$$\vdots$$

$R_{t1}^c, \ldots, R_{tm_t}^c$ denote the circles given by the intersection of Z with $R^s \in I_t$

and let $n_p = m_1 + m_2 + \ldots + m_t$.

Form histogram estimates for $g(x)$ and $f(x)$, respectively as follows.

$$g_{n_0}(x) = \begin{cases} \frac{1}{n_0 \triangle r} & \sharp \text{ spheres with radii in } I_j \qquad \text{x} \subset I_j \\ = 0 & \text{otherwise} \end{cases}$$

$$g_{n_0}(x) = \begin{cases} \frac{1}{n_0 \triangle r} \sharp \{R_I^s \in I_j\} & \text{x} \in I_j \\ 0 & \text{otherwise} \end{cases}$$

Similarly define

$$f_{n_p}(x) = \begin{cases} \frac{1}{n_p \triangle r} \{R_i^c \in I_j\} & \text{x} \in I_j \\ 0 & \text{otherwise} \end{cases}$$

Then $g_{n_0}(x)$ and $f_{n_p}(x)$ are histogram estimates of the densities of $G(r)$ and $F(r)$, respectively. One can write

$$f_{n_p}(x) = \frac{1}{n_p \triangle r} \sum_{j=1}^{t} \sum_{i=1}^{t} \sum_{k=1}^{m_i} I[R_{ik}^c \in I_j]I[x \in I_j]$$

Taking expectations and letting $R_{k(i)}^s$ be the spherical radius associated with R_{ik}^c. Let $\mathcal{F}$ the sigma field generated by the random variables $R_{ik}^c, R_{ik}^s, i, k \geq 1$.

$$\begin{aligned} E[f_{n_p}(x) \mid \mathcal{F}] &= \sum_{j=1}^{t} \sum_{i=1}^{t} \sum_{k=1}^{m_i} E[\frac{1}{n_p \triangle r}]P(R_{ik}^c \in I_j)I[x \in I_j] \\ &= \sum_{j=1}^{t} \sum_{i=1}^{t} \sum_{k=1}^{m_i} E[\frac{1}{n_p}\frac{T(x_{j-1} \mid R_{k(i)}^s) - T(x_j \mid R_{k(i)}^s)}{\triangle r}]I[x \in I_j] \\ &= \sum_{j=1}^{t} \sum_{i \geq j-1} \sum_{k=1}^{m_i} E[\frac{1}{n_p}]\frac{T(x_{j-1}, R_{k(i)}^s) - T(x_j, R_{k(i)}^s)}{\triangle r}I[x \in I_j] \\ &= \sum_{j=1}^{t} \sum_{i \geq j-1} E[\frac{1}{n_p}]\frac{x}{R_{k(i)}^s \sqrt{(R_{k(i)}^s)^2 - x^2}}I[x \in I_j] + o(1) \end{aligned}$$

9

Observe that $f(y) = \frac{x}{y\sqrt{y^2-x^2}}$ is decreasing in y. Giving

$$
\begin{aligned}
E[f_{n_p}(x) \mid \mathcal{F}] \;\leq\; & \sum_{j=1}^{t} \sum_{i \geq j-1} \sum_{k=1}^{m_i} E[\frac{1}{n_p}] \frac{x}{x_i\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \sum_{j=1}^{t} \sum_{i \geq j-1} E[\frac{m_i}{n_p}] \frac{x}{x_i\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \sum_{j=1}^{t} \sum_{i \geq j-1} E[\frac{m_i}{l^2} \frac{l^2}{n_p}] \frac{x}{x_i\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{l^2}{n_p} \sum_{j=1}^{t} \sum_{i \geq j-1} E[\frac{m_i}{l^2}] \frac{x}{x_i\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
\leq\; & \frac{l^2}{n_p} \sum_{j=1}^{t} \sum_{i \geq j-1} N_{v_i} 2x_i \frac{x}{x_i\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{l^2}{n_p} \sum_{j=1}^{t} \sum_{i \geq j-1} \frac{2n_{0i}}{l^3} \frac{x}{\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{1}{N_v 2\bar{R}} \sum_{j=1}^{t} \sum_{i \geq j-1} \frac{n_{0i}}{l^3} \frac{x}{\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{1}{N_v 2\bar{R}} \sum_{j=1}^{t} \sum_{i \geq j-1} \frac{n_{0i}}{n_0} \frac{n_0}{l^3} \frac{x}{\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{1}{\bar{R}} \sum_{j=1}^{t} \sum_{i \geq j-1} \frac{n_{0i}}{n_0} \frac{x}{\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{1}{\bar{R}} \sum_{j=1}^{t} \sum_{i \geq j-1} \frac{n_{0i}}{n_0 \, \triangle r} \frac{x}{\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{1}{\bar{R}} \sum_{j=1}^{t} \sum_{i \geq j-1} g_{n_0}(x_i) \frac{x}{\sqrt{x_i^2 - x^2}} I[x \in I_j] + o(1) \\
=\; & \frac{1}{\bar{R}} \sum_{j=1}^{t} g_{n_0}(x_i) \frac{x}{\sqrt{x_i^2 - x^2}} \sum_{j=1}^{i} I[x \in I_j] \triangle r + o(1) \\
\rightarrow\; & \frac{1}{\mu} \int_{x}^{\infty} \frac{x}{\sqrt{y^2 - x^2}} g(y) dy
\end{aligned}
$$

10

as $l \to \infty$ and $t \to \infty$, by the strong law of large numbers. So we have the inequality

$$f(x) = \lim_{l \to \infty} \lim_{t \to \infty} f_{n_p} \leq \frac{1}{\mu} \int_x^\infty \frac{x}{\sqrt{y^2 - x^2}} g(y) dy.$$

Similarly , a $\geq$ inequality holds

4 The Abel Integral Equation

In the previous section we observed that an integral relationship exists between $f(x)$ and $g(x)$ the densities of the circular radii and the spherical radii, respectively, of stereology, viz

$$f(y) = \frac{y}{m} \int_y^\infty \frac{g(x)}{\sqrt{x^2 - y^2}} dx \quad (2.1)$$

where $m = \int_0^\infty x g(x) dx$ is the mean spherical radius, and $R_{\max} = \infty$. By inversion

$$\begin{aligned} g(x) &= \frac{-2xm}{\pi} \int_x^\infty \frac{1}{\sqrt{y^2 - x^2}} \frac{d}{dy} [\frac{f(y)}{y}] dy \quad (2.2) \\ &= \frac{12m}{\pi} \frac{d}{dx} \int_x^\infty \frac{1}{\sqrt{y^2 - x^2}} f(dy) \end{aligned}$$

see Hochstat, page 41.

The problem is to use equation (2.1) to estimate $g(x)$ based on the observed data $y_1, y_2, \ldots, y_n$, the radii of the n circles on the test plane. There are several ways to use the data and the integral equation (2.1) to get an estimate of $f(x)$. Perhaps the simplest approach is the parametric, where it is assumed that

$$g(x) = g(x, \theta) \qquad \theta \in R^k$$

where g is known completely except for θ. For example, it may be assumed that $g(x, \theta)$ is either normal, lognormal, gamma, etc. The problem is reduced to estimating the parameters and using (2.2). As an example Dehoff(1965) used a lognormal distribution.

$$g(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\{-\frac{1}{2}(\frac{\ln(x) - \mu}{\sigma})^2\} \qquad x > 0$$

11

Another approach is the nonparametric approach, where the densities $f(x)$ and $g(x)$ are estimated by histograms and the integral is approximated by a sum. This method is called the finite difference or method of moments approach.

Divide (0,a] into subintervals (histogram bins) of width Δr_i and let $I_1 = (0, r_1], I_2 = (r_1, r_2], \ldots, I_k = (r_{k-1}, r_k]$ denote the bins. let

$$
\begin{aligned}
f_i & = \frac{1}{\Delta r_i}\{\text{number of} \quad y_j \in I_i\} \\
& \approx \int_{r_{i-1}}^{r_i} f(y)dy
\end{aligned}
$$

Now use any quadrature approximation on the integral in (2.1) i.e.

$$
\frac{y}{m}\int_y^\infty \frac{g(x)}{\sqrt{x^2 - y^2}}dx \approx \sum_{j=1}^K a_{ij}g_j
$$

$i = 1, 2, \ldots, k$, where K is the number of intervals in the quadrature approximation, a_{ij} is the quadrature weights and f_j is the value of g at the nodes. We get a system of equations.

$$
\begin{aligned}
f_1 & = \sum_{j=1}^K a_{1j}g_j \\
f_2 & = \sum_{j=1}^K a_{2j}g_j \\
& \vdots \\
f_k & = \sum_{j=1}^K a_{kj}g_j
\end{aligned}
$$

These can be written in matrix form as

$$
\mathbf{f} = A\mathbf{g}
$$

where

$$
\begin{bmatrix} f_1 \\ f_2 \\ f_k \end{bmatrix} = A \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_K \end{bmatrix}
$$

12

where $A = (a_{ij})$ is the kxK matrix of quadrature weights. For example consider the approximation where one takes

$$g(x) = \sum_{j=1}^{K} g_j I[x = R_j]$$

where $R_1, R_2, \ldots, R_K$ are a sequence of real numbers, and then g_j's are a sequence of constants. One may assume that the R_j's are a sequence of discretized values that the spherical radii can take on and the g_j's are the unknown frequencies of spherical radii taking on the value R_j. This is the Saltikov(1967) model where he assumes that the spherical radii are discretized into K species $R_1, \ldots, R_K$. Substituting into the integral in (2.1) gives

$$\int_y^a \frac{g(x)}{\sqrt{x^2 - y^2}} = \int_y^a \frac{1}{\sqrt{x^2 - y^2}} \sum_{j=1}^{K} g_j I[x = R_j] dx$$

$$= \sum_{j=1}^{K} g_j \int_y^a \frac{1}{\sqrt{x^2 - y^2}} I[x = R_j] dx$$

$$= \sum_{j=1}^{K} \frac{g_j}{\sqrt{R_j^2 - y^2}} \int_y^a I[x = R_j] dx$$

$$= \sum_{j=1}^{K} \frac{g_j}{\sqrt{R_j^2 - y^2}} I[y \leq R_j < \infty]$$

so that

$$f(y) = \sum_{j=1}^{K} \frac{g_j}{\sqrt{R_j^2 - y^2}} I[y \leq R_j < \infty]$$

Recall,

$$f_i = \int_{r_{i-1}}^{r_i} f(y) dy$$

$$= \int_{r_{i-1}}^{r_i} \frac{y}{m} \sum_{j=1}^{K} \frac{g_j}{\sqrt{R_j^2 - y^2}} I[y \leq R_j < \infty] dy$$

$$= \sum_{j=1}^{K} \frac{g_j}{m} \int_{r_{i-1}}^{r_i} \frac{y}{\sqrt{R_j^2 - y^2}} I[y \leq R_j < \infty] dy$$

13

$$= \sum_{j=1}^{K} \frac{g_j}{m}(1/2) \int_{r_{i-1}^2}^{r_i^2} \frac{1}{\sqrt{R_j^2 - y}} I[y \leq \infty] dy$$

$$= \sum_{j=1}^{K} \frac{g_j}{m} \int_{R_j^2-r_i^2}^{R_j^2-r_i^2} \frac{1}{\sqrt{t}}$$

$$= \sum_{j=1}^{K} \frac{g_j}{m} [\sqrt{R_j^2 - r_i^2} - \sqrt{R_j^2 - r_{i-1}^2}]$$

So that,

$$a_{ij} = \sqrt{R_j^2 - r_i^2} - \sqrt{R_j^2 - r_{i-1}^2}$$

The matrix A is upper-triangular, ie

$$a_{ij} = 0 \quad \text{if } r_i > R_j$$

So, if $K = k$ it is very simple to solve the matrix equation by successive elim ination, starting with the largest spheres. It is often convenient to choose $K = k$ then the solution is unique.

Other approximations can also be considered. For example, one may partition the range of g into subintervals, $(x_{i-1}, x_i], i = 1, \ldots K$, and in each subinterval approximate g with the polynomial

$$p_j(y) = \sum_{j=0}^{K} a_{ij} y^j \quad x_{j-1} \leq y < x_i$$

the coefficients depending on the data.

The histogram methods require grouped data. A way to avoid this is to rewrite the integral equation (2.1) in terms of the cumulative distribution function. By multiplying (2.1) by $(x^2 - z^2)^{-1/2}$ and integrating x from z to infinity one gets

$$1 - G(z) = \frac{2\mu}{\pi} \int_z^\infty \frac{F(dx)}{\sqrt{x^2 - z^2}} \qquad (2.3)$$

Note that by letting $z \to 0$ gives

$$1 = \frac{2\mu}{\pi} \int_0^\infty \frac{F(dx)}{x}$$

14

a method to estimate μ.

Let $F_n(x)$ denote the empirical distribution function

$$
\begin{aligned}
F_n(x) &= \frac{1}{n}\sharp\{y_i \leq x\} \\
&= \frac{1}{n}\sum_{i=1}^{n} I[y_i \leq x]
\end{aligned}
$$

Then substituting $F_n(x)$ for $F(x)$ in (2.3) gives the estimate

$$
1 - \hat{G}(x) = \frac{2\mu}{\pi}\sum_{y_i > x}\frac{1}{n}(y_i^2 - x^2)^{-1/2}
$$

This appears to be a good estimator but suffers several defects, for example $\hat{G}$ estimates a nondecreasing function, but

$$
\lim_{x \to y_i}\hat{G}(x) = -\infty
$$

Another nonparameteric method is the product-integration technique. Anderssen and Jakeman(1975,1975a) published several papers on its use. The technique is used to circumvent the problem of the integrand in (2.1) having a singularity. The method smooths out the sample distribution function using a localized Lagrange interpolation and then the singularity is integrated out analytically.

5   Summary of Methods

The histogram(finite difference) methods provide poor estimates from a statistical and numerical perspective. Problems occur because the integrand has a singularity and because the data are noisy. Watson(1971) discusses the poor statistical properties (such as infinite mean squared error). Anderssen(1976) shows that if $\triangle$ is the histogram bin width then a pertubation in a profile frequency is locally amplified by a factor of the order $\frac{1}{\triangle}$ into an estimated sphere size frequency.

15

One way to overcome the noisy data problem is to use methods for solving the ill-posed inverse problem. Nychka et al. (1984) use these techniques to provide spline estimators.

Product integration and spectral differentiation appears to be mathematically superior to finite difference methods. But, for practical purposes they are mathematically complex, especially spectral integration.

Of course the problem with the parametric approach is having to commit oneself to a precise form of $g(x)$, especially in those situations, very common in biology, in which $g(x)$ may be a mixture of an unknown number of components. Furthermore, one needs good parameter estimation techniques.

Most of the comments in this section are taken from the paper of Cruz-Orive (1982).

References

[1] Anderssen R.S.,and Bloomfield (1974a) A time series approach to numerical differentiation. *Technometrics 16*, 69.

[2] Anderssen and Bloomfield (1974b) Numerical differentiation procedures for non-exact data. *Num. Math 22* 157.

[3] Anderssen, Jakeman (1975) Abel type integral equations in stereology II. *Journal of Microscopy 105*, 135.

[4] Anderssen and Jakeman (1975) Product integration for functionals of particle size distributions *Util. Math.*

[5] Cruz-Orive L.M. (1982) Distribution-free estimation of sphere size distributions from slabs showing overprojection and truncation, with a review of previous methods.*Journal of Microscopy 131* 265.

[6] DeHoff R.T.(1965) The estimation of particle size distributions from simple counting measurements made on random plane sections. *Trans. Metall. Soc. A.I.M.E. 233*, 25.

[7] Hochstadt, H. *Integral Equations* Wiley 1973.

[8] Keiding N., Jensen S.T., and Ranek L. (1972) Maximum likelihood estimation of the size distribution of liver cell nuclei from the observed distribution in a plane section. *Biometrics 28* 813.

[9] Nychka D., Wahba G., Goldfarb S. and Pugh T. (1984) Cross-validated spline methods for estimation of three dimemsional tumor size distributions from observations on two-dimensional cross sections. *JASA 79* 832.

[10] Saltikov S.A. (1967) The determination of the size distribution of particles in an opaque material from a measurement of the size distribution of their sections. In *Stereology ,ed. by H. Elias* 163-73. Springer Inc. New York.

[11] Watson G.S. (1971) Estimating functionals of particle size distributions *Biometika 58* 483.

[12] Wicksell S.D. (1925) The corpuscle problem *Biometrika 17* 84.

[13] Wicksell S.D.(1926) The corpuscle problem II. *Biometrika 18* 151.

NIST-114A (REV. 3-90)	U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY **BIBLIOGRAPHIC DATA SHEET**	1. PUBLICATION OR REPORT NUMBER NISTIR 4370
		2. PERFORMING ORGANIZATION REPORT NUMBER
		3. PUBLICATION DATE JULY 1990

4. TITLE AND SUBTITLE

A Mathematical Treatment of the Spherical Stereology

5. AUTHOR(S)

Charles Hagwood

6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS) U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY GAITHERSBURG, MD 20899	7. CONTRACT/GRANT NUMBER
	8. TYPE OF REPORT AND PERIOD COVERED

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)

10. SUPPLEMENTARY NOTES

11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)

Stereology is the study of 3-dimensional particles based on their two dimensional cross-sections. In this paper the particles are assumed to be spheres with a random radii distribution. We considered the mathematical aspects of this type of stereology.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

density estimation; inverse problem; probability; spheres; stereology; unfolding

13. AVAILABILITY	14. NUMBER OF PRINTED PAGES
[X] UNLIMITED [ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS). [ ] ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402. [X] ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.	21
	15. PRICE A02

ELECTRONIC FORM